

An Uncertainty-Aware Approach for Exploratory Microblog Retrieval

Mengchen Liu, Shixia Liu, Xizhou Zhu, Qinying Liao, Furu Wei, and Shimei Pan

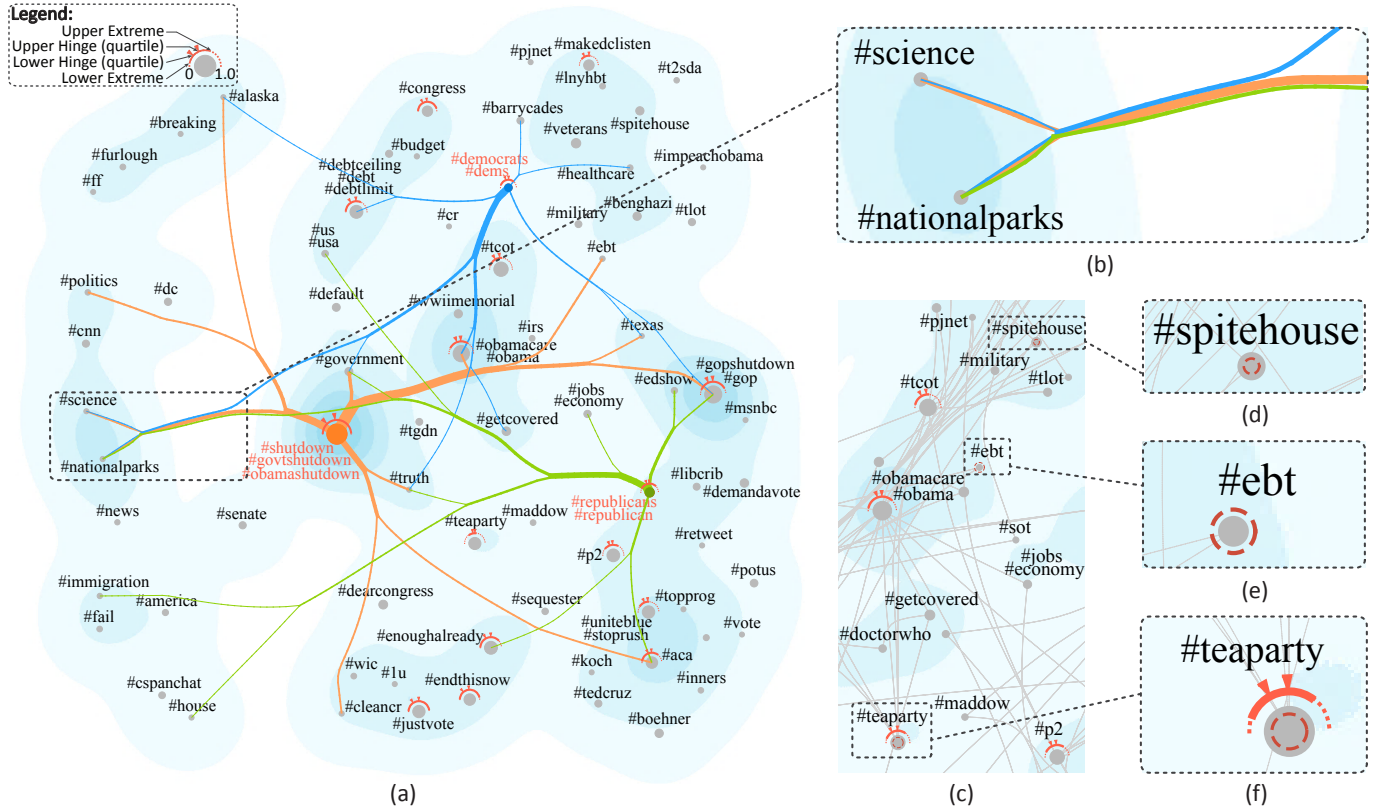


Fig. 1. Exploratory retrieval of the government shutdown dataset: (a) the hashtag graph with uncertainty and its propagation; (b) uncertainty propagation; (c)-(f) interactive ranking refinement results.

Abstract— Although there has been a great deal of interest in analyzing customer opinions and breaking news in microblogs, progress has been hampered by the lack of an effective mechanism to discover and retrieve data of interest from microblogs. To address this problem, we have developed an uncertainty-aware visual analytics approach to retrieve salient posts, users, and hashtags. We extend an existing ranking technique to compute a multifaceted retrieval result: the mutual reinforcement rank of a graph node, the uncertainty of each rank, and the propagation of uncertainty among different graph nodes. To illustrate the three facets, we have also designed a composite visualization with three visual components: a graph visualization, an uncertainty glyph, and a flow map. The graph visualization with glyphs, the flow map, and the uncertainty analysis together enable analysts to effectively find the most uncertain results and interactively refine them. We have applied our approach to several Twitter datasets. Qualitative evaluation and two real-world case studies demonstrate the promise of our approach for retrieving high-quality microblog data.

Index Terms—microblog data, mutual reinforcement model, uncertainty modeling, uncertainty visualization, uncertainty propagation.

1 INTRODUCTION

Microblogs such as Twitter and Facebook are among the most popular platforms for people to share their daily observations and thoughts,

including personal status updates and opinions regarding products or government policies. Since the crowd in microblogs provides many individual comments/opinions that were not available before, businesses and organizations have begun to leverage microblogs to profile customers, derive brand perception, gauge citizen sentiments, and predict the stock market [23, 34, 41, 53]. For example, retailers track and examine relevant microblog posts to understand customer opinion toward their products and services. In spite of the growing interest in quickly analyzing customer opinions or breaking news in microblogs, progress has been hampered by the lack of an effective mechanism to retrieve data of interest from microblogs.

For this reason, researchers have developed a number of microblog retrieval methods [6, 17]. The main goal is to generate a list of k microblog posts that are relevant to the information needs represented by a query q . Although these methods have successfully retrieved

- M. Liu and S. Liu are with Tsinghua University. E-mail: simon900314@139.com, shixia@tsinghua.edu.cn. S. Liu is the corresponding author.
- X. Zhu is with USTC. E-mail: ezra0408@mail.ustc.edu.cn.
- Q. Liao and F. Wei are with Microsoft. E-mail: {qiliao, fuwei}@microsoft.com.
- S. Pan is with University of Maryland, Baltimore County. E-mail: shimei@umbc.edu

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication xx Aug. 2015; date of current version 25 Oct. 2015.

For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

relevant posts, they have two major drawbacks. First, the unique characteristics of microblog data are not comprehensively considered to improve retrieval performance. Posts, users, and hashtags are three key dimensions of microblog data. These dimensions are not independent, as they often influence one another. For example, a post published by an influential user and labeled with a popular hashtag tends to be salient. However, the existing approaches to microblog retrieval do not tightly integrate the three dimensions and do not take advantage of the relationships among them. These approaches either consider only the post or treat it as the primary dimension and the others as secondary dimensions to filter posts. For example, ScatterBlogs2 [6] finds posts of interest by checking whether the posts contain a certain hashtag. Second, existing methods do not address uncertainty in the retrieval models. Improving the modeling and presentation of uncertainty can help describe retrieved data more accurately, which can in turn assist analysts in making more informed decisions [14, 39, 49].

To address the above problems, we have developed an uncertainty-aware microblog retrieval toolkit, MutualRanker, to estimate uncertainty introduced by the analysis algorithm as well as to quickly retrieve salient posts, users, and hashtags. Since posts are shared and propagated on social networks, the authority of an author and the popularity of a hashtag play an important role in determining the importance of a post and vice versa. Accordingly, we formulate uncertainty-aware microblog retrieval using an uncertainty-based mutual reinforcement graph model (MRG) [16, 45], where the content quality of posts, the social influence of users, and the popularity of hashtags mutually reinforce one another. We adopt a Monte Carlo sampling method to solve MRG because of its effective local update mechanism, fast convergence, and probability-based uncertainty formalization [2]. A dispersion-based measure is used to estimate the uncertainty generated by the Monte Carlo sampling method. In addition, we model the uncertainty propagation as a Markov chain.

To help analysts understand the retrieved data, we have designed a composite visualization [20]. Specifically, a density-based graph visualization has been developed to visually illustrate posts, users, hashtags, and their relationships. An uncertainty glyph and a flow map are employed to represent uncertainty and its propagation on a graph (Fig. 1). The three visualization components, together with the uncertainty analysis, enable analysts to quickly detect the most uncertain results and interactively resolve them. The Monte Carlo sampling method is then used to incrementally modify the ranking results to meet user needs.

In summary, our work presents three technical contributions:

- **An uncertain-aware microblog retrieval model** that extracts salient posts, users, and hashtags. This model also computes the associated uncertainty and its propagation among graph nodes.
- **A composite visualization** that enables users to understand the three-level, mutual reinforcement ranking results, the associated uncertainty, and uncertainty propagation patterns.
- **A visual analytics system** that helps users quickly retrieve data of interest, as well as analyze and understand the ranking results in an interactive and iterative process.

2 RELATED WORK

2.1 Microblog Retrieval

In the field of data mining, a number of approaches have been proposed to retrieve data from microblogs. A comprehensive survey was presented by Cherichi and Faiz [12]. Most recent work can be categorized into two groups: vector-space-based approaches and link analysis approaches.

The vector-space-based approach employs two feature vectors to represent a query and a post. A similarity measure (e.g., cosine similarity) is then adopted to estimate the similarity between the post and the query. There have been some recent research efforts that exploit additional structural features such as URLs and hashtags to enhance retrieval performance [1, 29, 31].

Recently, to take advantage of the link structure of social networks, researchers have introduced the PageRank algorithm [7] in microblog retrieval. For example, TwitterRank [46] adopts the follower-follower

link structure and the PageRank algorithm to identify influential users. Duan et al. [16] modeled the tweet-ranking problem as an MRG [45], where the social influence of users and the content quality of tweets mutually reinforce each other. Specifically, the post graph, the user graph, and the hashtag graph, as well as the relationships between the three graphs, were used to retrieve salient posts, users, and hashtags. We extend this approach by explicitly modeling the uncertainty of the ranking result, as well as its propagation on the tweet/user/hashtag graph.

In the field of visual analytics, a great deal of research has been conducted on visually analyzing microblog data. The methods applied include event detection [30], topic extraction and analysis [25, 40, 50], information diffusion [8, 52], sentiment analysis [47, 48], and revenue/stock prediction [28, 34]. However, few studies have focused on microblog retrieval.

Bosch et al. [6] developed ScatterBlogs2 to extract microblog posts of interest. It allows analysts to build customized post filters and classifiers interactively. These filters and classifiers are then utilized to support real-time post monitoring. In post filtering, the post dimension is considered the primary dimension and the hashtag the secondary dimension. In contrast, we tightly integrate the posts, users, and hashtags in the MRG model and use the model to retrieve high-quality microblog data. Moreover, we also model uncertainty in the retrieval process. Since analysts can interactively refine the model, we can further improve retrieval quality by leveraging the uncertainty formalization and analysts' knowledge.

2.2 Interactive Uncertainty Analytics

Frequently, uncertainty is introduced into visual analytics when data is acquired, transformed, or visualized [14, 24, 27]. A number of uncertainty analysis methods have been proposed, which can be categorized into two groups: uncertainty visualization and uncertainty modeling.

Many studies on uncertainty visualization have been conducted in the field of geographic visualization and scientific visualization [32, 37, 42]. Typical uncertainty representation techniques include the addition of glyphs and geometry, the modification of geometry and attributes, animation, sonification, and psycho-visual approaches [32]. Recently, researchers are increasingly interested in the design of uncertainty representations for information visualization and visual analytics. For example, Collins et al. [13] designed two alternatives, the gradient border and the bubble border, to illustrate uncertainty in lattice graphs. Wu et al. [48] developed a circular wheel representation and subjective logic to convey uncertainty in customer review analysis. Slingsby et al. [38] utilized bar charts to reveal the uncertainty associated with geodemographic classifiers. To represent uncertainty in aggregated vertex sets, Vehlow et al. [43] considered the lightness and shape of the node. Chen et al. [10] adopted the uncertainty histogram to explore uncertainty in the context of a multidimensional ensemble dataset. Compared with these methods, MutualRanker not only visualizes uncertainty, but also its propagation on a graph. We also support users to interactively modify the uncertain result.

Another type of uncertainty visualization represents the uncertainty in the analysis process. Zuk and Carpendale [55] studied issues related to uncertainty in reasoning and determined the type of visual support required. Correa et al. [14] developed a framework to represent and quantify the uncertainty in the visual analytics process. Wu et al. [49] extended this framework to show the uncertainty flow in the analysis process. By contrast, our work aims to model uncertainty in microblog retrieval. We focus on visually illustrating topological uncertainty propagation on a graph and on designing an iterative visual analytics process to actively engage analysts in reducing overall uncertainty.

Probability theory, fuzzy set theory, rough set theory, and evidence theory are four major approaches to model uncertainty [54]. Among these approaches, probability theory is the most commonly used method in visual analytics. For example, Correa [14] and Wu et al. [49] regarded uncertainty as a parameter that describes the dispersion of measured values. Specifically, they represented uncertainty as an estimated standard deviation, in which the measured value is defined on the set of both positive and negative real numbers. Since the measured value (the ranking score) in our approach is defined on the set of positive real

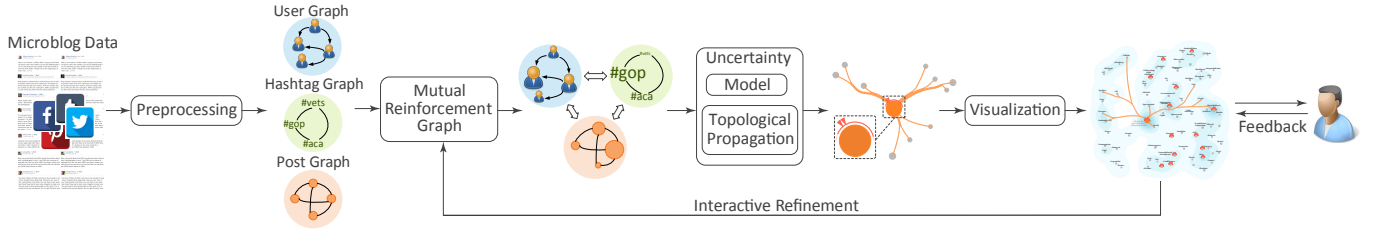


Fig. 2. MutualRanker pipeline.

numbers, the above modeling method cannot be directly applied to our work. Therefore, we employ a Poisson mixture to model uncertainty.

3 MUTUALRANKER OVERVIEW

3.1 Requirement Analysis

The research problems were gradually identified in our own research projects related to Twitter data analysis. In these projects, we often needed to discover and retrieve relevant tweets, users, and hashtags by keyword search. Frequently, we also needed to manually check the data and improve the quality using heuristics. This process can be very time-consuming and requires domain expertise.

To address this issue, we collaborated with two domain experts to develop MutualRanker, including one researcher in sociology (S) and one researcher in media and communications (C). The experts are experienced in retrieving data from microblogs. They also had experience using a method similar to the one described above. We conducted several interviews with them, mainly focusing on probing their needs and microblog retrieval process. We then identified the following high-level requirements based on their feedback.

R1 - Examining an initial set of salient microblog data. Both experts expressed the need for a ranking list of keyword search results. Keyword-based microblog retrieval results often include millions of posts and tens of thousands of users and hashtags. Thus, these results are too massive for analysts to quickly discover relevant data. The experts usually have to examine the data carefully and design a set of rules to filter out irrelevant data. As a result, they stated the need for a toolkit that can rank extracted posts, users, and hashtags to facilitate their data retrieval tasks. This need is consistent with the findings of previous research [16, 46].

R2 - Revealing relationships within microblog data. Previous research [11, 16, 25] has also indicated that the relationships within data help users locate interesting information more easily. Furthermore, the relationships among the three dimensions of microblog data (posts, users, and hashtags) can assist them in extracting salient data. For example, posts from opinion leaders are usually more important than those from average users. The domain experts desired the ability to explore different types of relationships.

R3 - Exploring salient microblog data from different perspectives. Since the three dimensions of microblog data usually influence each other, the experts wanted to understand this influence so that they can link important data in one dimension to that in another dimension. For example, expert S said that, “Collecting relevant tweets is very important for some of our projects. After finding one important tweet, I usually check other tweets from the same author as well as the tweets marked by the same hashtag(s). This helps me find relevant tweets quickly.”

R4 - Understanding the error produced by the ranking mechanism. The microblog data ranking mechanism is not perfect and often introduces errors or uncertainty into the retrieval process. Thus, the degree of uncertainty must be analyzed and understood to facilitate informed decision-making [14, 37, 49]. The experts requested to know which ranking scores are more error-prone.

R5 - Analyzing the influence of the errors of one item on other items. The experts also expressed the need to understand error propagation among data items. They claimed that this information can help them considerably in filtering out irrelevant data. For example, expert C commented, “When I find an item with an incorrect ranking

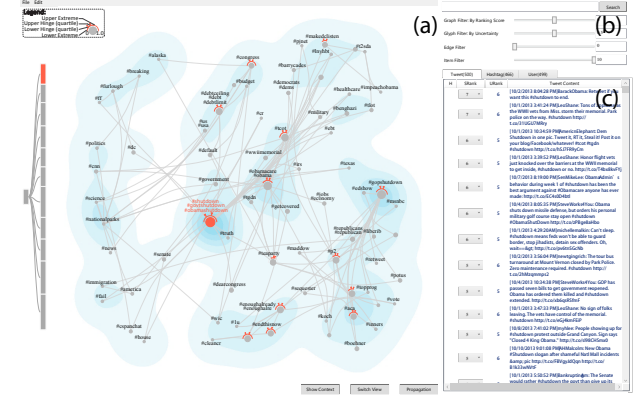


Fig. 3. User interface: (a) MutualRanker visualization; (b) control panel; (c) information panel.

score, I also want to know which items are influenced by this so that I can adjust the ranking score quickly.”

3.2 System Overview

The collected requirements have motivated us to develop a visual analytics toolkit, MutualRanker. It consists of the following components:

- An MRG model to generate the initial ranking lists of posts, users, and hashtags (**R1**);
- An uncertainty model to estimate uncertainty and its topological propagation on a graph (**R4, R5**);
- A composite visualization to present the graph-based ranking results, uncertainty, and its propagation (**R2, R3**).

The primary goal of MutualRanker is to extract a list of k microblog posts/users/hashtags that are relevant to query q . Fig. 2 illustrates the main components needed to achieve this goal. Given a microblog dataset extracted by a query, the preprocessing module first extracts the post graph, the user graph, and the hashtag graph. The three graphs are then fed to the MRG model, which produces three ranking lists of posts, users, and hashtags. The uncertainty module estimates the uncertainty in the retrieval model and its topological propagation. The visualization module takes the ranking results and the uncertainty estimation as input and illustrates them in a composite visualization that includes a graph visualization, an uncertainty glyph, and a flow map. Users can interact with the generated visualization for further analysis. For example, a user can modify a ranking result. With this input, MutualRanker will incrementally update the ranking results.

Fig. 3 depicts the user interface of MutualRanker. It contains three different interaction areas: MutualRanker visualization (Fig. 3(a)), control panel (Fig. 3(b)), and information panel (Fig. 3(c)). The visualization view consists of two parts: 1) the stacked tree visualization that shows the hierarchical structure of microblog data; 2) the composite visualization that simultaneously reveal the retrieved microblog data, the uncertainty of the ranking results, and its topological propagation. The control panel consists of a set of controls that enable users to interactively update the ranking. The information panel displays the corresponding microblog data such as posts, users, and hashtags for a selected aggregate item.

4 MUTUAL REINFORCEMENT GRAPH

The main feature of MRG [16, 45] is that it employs both the relationships within posts, users, or hashtags, and the relationships between them to improve rankings. This feature significantly reduces the workload of analysts when interacting with our visual analytics system. For example, if an analyst modifies the ranking score of a hashtag, MRG not only incrementally updates the ranking scores of the neighboring hashtags, but also those of relevant users and posts. This process allows our system to integrate user knowledge into the visual analytics process with acceptable user effort. This is also the main reason why we adopt MRG in MutualRanker.

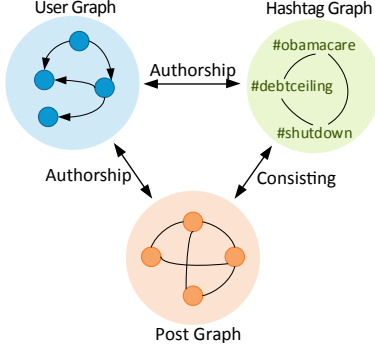


Fig. 4. Mutual reinforcement model.

The input of MRG includes three graphs, the post graph, the user graph, and the hashtag graph, as well as the relationships among them. The three graphs and their relationships are shown in Fig. 4. As in [16], the post graph is built based on cosine similarity. A recent study has shown that cosine similarity with a TFIDF weighting scheme is the most appropriate measure to compute the similarity between microblog posts [35, 51]. As a result, we employ cosine similarity in our system. The user graph is constructed based on follower-followee relationships. The hashtag graph is generated according to the co-occurrence of two hashtags. The three graphs are also connected by two relationships: authorship and co-occurrence. If a user publishes a post, then we connect this user with his/her post. We also link this user with all of the hashtags in this post. Each post is also linked to all of the hashtags associated with it.

For simplicity, we uniformly denote posts, users, and hashtags as items in the following discussion.

The MRG employs a method similar to PageRank [7] to model the mutual influence among different items in heterogeneous graphs:

$$\begin{bmatrix} R_p \\ R_u \\ R_h \end{bmatrix} = d \begin{bmatrix} \alpha_{pp}M_{pp} & \alpha_{up}M_{up} & \alpha_{hp}M_{hp} \\ \alpha_{pu}M_{pu} & \alpha_{uu}M_{uu} & \alpha_{hu}M_{hu} \\ \alpha_{ph}M_{ph} & \alpha_{uh}M_{uh} & \alpha_{hh}M_{hh} \end{bmatrix} \begin{bmatrix} R_p \\ R_u \\ R_h \end{bmatrix} + (1-d) \begin{bmatrix} W_p \\ W_u \\ W_h \end{bmatrix}. \quad (1)$$

R_p , R_u , and R_h are the ranking score vectors of posts (p), users (u), and hashtags (h). M_{xy} denotes the affinity matrix from x to y , where x, y can be posts, users, or hashtags. α_{xy} is a weight used to balance the mutual reinforcement strength among posts, users, and hashtags. d is the damping factor in PageRank, and we set it to 0.85, as in [7]. W_p , W_u , and W_h are vectors for the prior saliency of the items (e.g., the content quality of posts, the social influence of users, or the popularity of hashtags).

$$\text{Let } R = \begin{bmatrix} R_p \\ R_u \\ R_h \end{bmatrix}, W = \begin{bmatrix} W_p \\ W_u \\ W_h \end{bmatrix}, \text{ and } M = \begin{bmatrix} \alpha_{pp}M_{pp} & \alpha_{up}M_{up} & \alpha_{hp}M_{hp} \\ \alpha_{pu}M_{pu} & \alpha_{uu}M_{uu} & \alpha_{hu}M_{hu} \\ \alpha_{ph}M_{ph} & \alpha_{uh}M_{uh} & \alpha_{hh}M_{hh} \end{bmatrix}.$$

Then, Eq. (1) can be simplified as:

$$R = dMR + (1-d)W. \quad (2)$$

5 MRG-BASED UNCERTAINTY ANALYSIS

Since exact inference of MRG is very time-consuming on a large graph, we approximate it using a more efficient Monte Carlo sampling method. We also explicitly model the uncertainty associated with each item (e.g., a post, a user, or a hashtag), as well as its propagation on the graph.

5.1 MRG Computation with Monte Carlo Sampling Method

Duan et al. [16] proposed a matrix-based method to solve MRG, which iteratively updates the ranking scores using Eq. (1). The matrix-based method is a global one. An update to any item is achieved by running the method on the entire item set, which is very time-consuming. To address this problem, we use the Monte Carlo sampling method. The advantages of this sampling method over the matrix-based method are as follows [2]:

- Ranking scores are locally updated when the input changes locally;
 - The ranking scores of important items are accurately estimated after a few iterations;
 - The uncertainty of the ranking scores are modeled accurately because the Monte Carlo method calculates variance statistically.
- To employ this method, we first solve R as formulated by Eq. (2):

$$R = (1-d)(I-dM)^{-1}W = (1-d)\left(\sum_{k=0}^{\infty} d^k M^k\right)W. \quad (3)$$

We then perform a series of random walks for each item. A random walk may stop at each step with a probability of $1-d$. If the walk continues, then it proceeds to the next step according to the matrix M . Each element m_{ij} defines the transition probability from i to j .

Let $Z = \sum_{k=0}^{\infty} d^k M^k$. The ranking score of each item is:

$$r_j = (1-d) \sum_{i=1}^N w_i z_{ij}, \quad (4)$$

where the element z_{ij} in Z is the average number of times that a random walk starting from item i visits item j . We estimate z_{ij} by computing the empirical mean of a number of random walks.

Duan et al. [16] only consider the similarity of items in computing m_{ij} . Thus, high-ranking scores may be incorrectly assigned to users who publish many posts that do not receive attention. To address this, we consider the prior saliency of items in the sampling process. Specifically, the transition probability from m_{ij} is defined as $\text{similarity}(i, j) \cdot w_j$.

5.2 Uncertainty Modeling

In MutualRanker, we use an approximation method to solve MRG, which may introduce uncertainty into the retrieval results. It is therefore important to model uncertainty. Since we employ the Monte Carlo sampling method, the distribution of each ranking score is known. Hence, we can employ the probability theory to model uncertainty.

Uncertainty is defined as a parameter for depicting the dispersion of values that can be reasonably attributed to the measured value [5]. Traditional methods model the measured value as a normally distributed random variable [14, 49]. Variance [14] and standard deviation [49] are among the most commonly used measures to represent uncertainty wherein the measured value is defined on the set of both positive and negative real numbers.

The measured value (ranking score) in our approach is defined on the set of positive real numbers. Thus, the above modeling method cannot be applied directly to our work.

According to [2], z_{ij} in Eq. (4) has a Poisson distribution. The ranking score is the weighted sum of a series of z_{ij} . Hence, the ranking score is modeled as a Poisson mixture. For a Poisson mixture, the variance is approximately proportional to the mean. Hence, if we use variance to model uncertainty, the larger the ranking score, the more uncertain it is, but this is not always true.

Standard deviation is the square root of variance and has a similar problem. Consequently, variance and standard deviation are not good measures for depicting uncertainty in our model.

For such a distribution, a commonly used measure of dispersion is the variance-mean-ratio (VMR) [15]. The higher the VMR, the more dispersed the distribution. For item j , its VMR (u_j) can be defined as:

$$u_j = v_j / r_j, \quad (5)$$

where v_j is the distribution variance of the ranking score of item j . According to [2], v_j can be calculated as follows:

$$v_j = (1-d)^2 \sum_{i=1}^N w_i^2 v_{z_{ij}}. \quad (6)$$

where $v_{z_{ij}}$ is the variance of z_{ij} . Each z_{ij} obeys a Poisson distribution and its variance can be calculated from its expectation.

The massive number of items in the microblog data means we cannot place all of them on the screen. Hence, we aggregate similar items to form a cluster. The overall ranking score of a cluster r_c is defined as the sum of the ranking scores of its items [4]. The ranking scores are independent of each other and the overall variance of the cluster, v_c , is the sum of the variance of the ranking scores. Thus, the uncertainty of a cluster, u_c , can be calculated naturally by dividing v_c and r_c .

$$u_c = v_c / r_c = \sum_{j \in c} (r_j / r_c) (v_j / r_j) = \sum_{j \in c} k_j u_j. \quad (7)$$

Eq. (7) shows that u_c can be expressed by a weighted sum of the uncertainty of its items where each weight k_j is the ratio of the ranking scores of item j and cluster c . Thus, the uncertainty of a cluster is mainly determined by its important items.

5.3 Topological Uncertainty Propagation

If an analyst finds an incorrectly ranked item, he can modify it based on his knowledge. He can further track how the uncertainty propagates from one cluster to another to identify other affected items. To help an analyst track uncertainty, we explicitly model its topological propagation on the graph.

In MRG, the ranking score of an item can be expressed as a linear combination of ranking scores of related items. Hence, the variances of a ranking score can also be expressed as a linear combination of the variances of related ranking scores. The uncertainty of each item can be calculated from its ranking score and its variance, and hence, the uncertainty of an item can also be expressed linearly by the uncertainty of other items. Specifically,

$$u_j = \sum_{i=1, i \neq j}^N m_{ij}^* u_i, \quad (8)$$

where each $m_{ij}^* = (d^2 m_{ij}^2 r_i) / ((1 - d^2 m_{jj}^2) r_j)$. Eq. (8) shows that the uncertainty of each item is not independent and it propagates on the graph in a linear form. Thus, for each pair of items i and j , $m_{ij}^* u_i$ can be viewed as the propagated uncertainty from item i to j . We denote it by $u_{i \rightarrow j}$.

Rewriting Eq. (8) in a matrix form, we can formulate the uncertainty propagation as a Markov chain:

$$U M^* = U, \quad (9)$$

where $U = [u_j]_{1 \times N}$ and $M^* = [m_{ij}^*]_{N \times N}$.

Similar to the uncertainty propagation from item to item, we can model the uncertainty propagation from cluster to cluster using the following procedure. First, based on Eq. (8), we calculate the propagated uncertainty from each item i in the source cluster c_s to each item j in the target cluster c_t (Fig. 5(a)). Second, for each item j in c_t , we compute the propagated uncertainty $u_{c_s \rightarrow j}$ from c_s to item j by aggregating the uncertainty propagated from each item in the source cluster (Fig. 5(b)).

$$u_{c_s \rightarrow j} = \sum_{i \in c_s} u_{i \rightarrow j}. \quad (10)$$

Finally, the uncertainty of a cluster is a weighted sum of the uncertainty of the items in it (Eq. (7)). Thus, the overall propagated uncertainty $u_{c_s \rightarrow c_t}$ from c_s to c_t can be calculated as the weighted sum of the propagated uncertainty from c_s to each j in c_t (Fig. 5(c)).

$$u_{c_s \rightarrow c_t} = \sum_{j \in c_t} k_j u_{c_s \rightarrow j}. \quad (11)$$

5.4 Incremental Ranking Update

We also allow analysts to interactively modify the item ranking result based on their knowledge. We can update the model locally because we use the Monte Carlo sampling method. After the analyst changes the ranking score(s), our approach iteratively updates the prior salience score(s) of the item(s). Accordingly, the affinity matrix is changed from

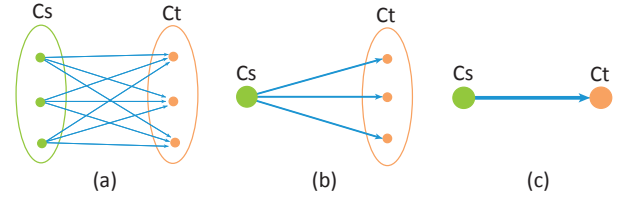


Fig. 5. Topological uncertainty propagation calculation

M to M' . This change only affects a small part of the random walks used in the Monte Carlo sampling method.

For the affected random walks, existing incremental graph ranking algorithms [3] perform re-sampling and update the ranking scores by aggregating the statistics of these new random walks into the original results. One main problem with these algorithms is that re-sampling requires a considerable amount of time, which may make real-time interaction impossible. Suppose n is the average number of neighbors that an item has and l is the average length of a sampled random walk. At each step in a random walk, we have to sample from a multinomial distribution with n possible outcomes and the time cost is $O(n)$. Thus, sampling a new random walk will take $O(nl)$ time. The time needed to compute and aggregate the statistics of these samples is $O(l)$. The total time required for a new sample is $O(nl) + O(l) = O(nl)$.

However, in our scenario, we do not delete or add edges on the graphs. As a result, we do not need to perform re-sampling. We only need to modify the statistics of a random walk based on the modified transition probability m_{ij} , thereby avoiding the high cost associated with resampling. The time cost of updating an influenced random walk is reduced to $O(l)$.

Given a random walk: $path = \{i \rightarrow n_1 \rightarrow \dots \rightarrow n_{k-1} \rightarrow j\}$, we define a new random variable x_{ij}^k . In particular, $x_{ij}^k = 1$ indicates that the random walk starts from i and reaches j by moving k steps. The original weight of each step in the random walk is 1. During an update, we re-calculate the weight of this step using $P'(x_{ij}^k = 1) / P(x_{ij}^k = 1)$. $P(x_{ij}^k = 1)$ is the probability of $x_{ij}^k = 1$ according to M and $P'(x_{ij}^k = 1)$ is the probability of $x_{ij}^k = 1$ according to M' . Hence, $P(x_{ij}^k = 1)$ is calculated by:

$$P(x_{ij}^k = 1) = m_{in_1} m_{n_1 n_2} \dots m_{n_{k-1} j}. \quad (12)$$

Similarly, $P'(x_{ij}^k = 1)$ can also be calculated.

6 VISUALIZATION

To help analysts extract microblog data of interest interactively, we have designed a composite visualization that includes a graph visualization, an uncertainty glyph, and a flow map (Fig 1(a)).

6.1 Ranking Results as Graph Visualization

Since one post corresponds to only one user and a few hashtags, the scope of influence of a post is smaller than that of a user or a hashtag. Updating the ranking score of a post will only directly affect the ranking scores of its author, a few related hashtags, and a number of posts. In contrast, updating the ranking score of a user or hashtag will directly impact the ranking scores of hundreds or even thousands of posts as well as a number of users and hashtags. On the other hand, the number of posts is usually huge, around 10-100 times that of users or hashtags. Analysts would require more time to provide their feedback on a post graph. As a result, we regard the user and the hashtag as the primary visualization elements and the post as a secondary element mainly used to illustrate the content of the primary elements. Accordingly, users and hashtags are visually represented by a node-link graph whereas posts are represented as a list. For simplicity, we take a hashtag graph as an example to illustrate the basic idea of graph visualization.

To allow analysts to navigate large graphs efficiently, a hierarchy is built based on a Bayesian Rose Tree [26] with each non-leaf node representing a hashtag cluster. As shown in Fig. 3(a), a stacked tree is adopted to represent the hashtag hierarchy and a density-based graph visualization is employed to illustrate the relationships within the user/hashtag graphs and between them (R2, R3).

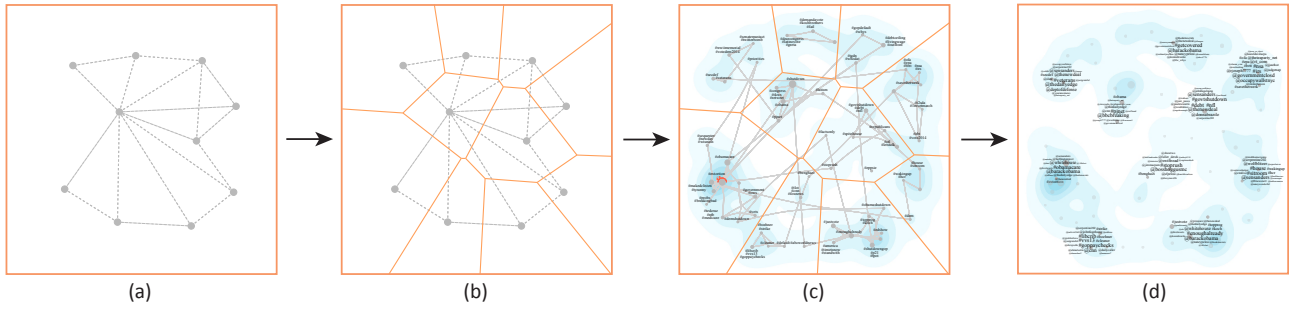


Fig. 6. Basic idea of the layout algorithm: (a) place the cluster nodes and derive the layout center of each cluster; (b) compute the Voronoi tessellation and treat the cell as the layout area of each cluster; (c) layout of representative and non-representative nodes; and (d) final layout result with context.

The density-based graph visualization combines a node-link diagram with a density map to display the nodes at the selected level of the hashtag tree. As in [25], we extract representative nodes for each of the cluster nodes at the selected tree level and assign other non-representative nodes to their closest representative nodes. As shown in Fig. 1(a), the representative nodes are displayed as a node-link diagram and the other nodes as a density map. In this visualization, the representative nodes of one cluster are placed near each other to reflect their closeness. The size of the node encodes the sum of the ranking score of each item. The corresponding users are overlaid around the selected hashtag node to provide more analysis context (Fig. 6(d)).

Layout. The layout of the stacked tree is quite straightforward. Thus, we introduce the layout of the density-based graph, which contains the following steps.

Step 1: Derive the layout center of each cluster at the selected tree level. We build a cluster graph by checking the edge connections between the two cluster nodes. An edge is added if a sufficient number of connections between the two cluster nodes can be found. The cluster graph is then placed by a force-directed layout [21]. As shown in Fig. 6 (a), the position of each cluster node is treated as the center of each hashtag cluster.

Step 2: Compute the layout area of each cluster. In this step, we compute the corresponding Voronoi tessellation based on the cluster center. The corresponding tessellation cells are treated as layout areas of the hashtag clusters (Fig. 6(b)).

Step 3: Layout of representative and non-representative nodes. In this step, the force-directed layout is adopted to place the representative nodes. To ensure the representative nodes within one cluster are placed in the corresponding cluster layout area, a repulsion force is added from the area boundary to each node within this area. The kernel density estimation [22] is utilized to represent the distribution of non-representative nodes (Fig. 6(c)).

Step 4: Layout of the context word cloud. Showing the hashtag graph and user graph simultaneously would introduce visual clutter. To solve this issue, we treat the hashtag graph as a primary element and the user information as context. In particular, when a hashtag node is selected, a word cloud that includes the users who use this hashtag is laid out to provide user context. In this word cloud, the selected hashtag is placed in the middle. A sweep-line-based word cloud layout algorithm [36] is employed to produce such a word cloud. Fig. 6 (d) shows a layout result with a word cloud context.

Interaction. The following interactions are provided to assist analysts in investigating the ranking results from multiple perspectives.

Examining the ranked microblog data and their relationships (R2). The density-based graph visualization provides an easy way to explore the ranking results from the hashtag or user perspective. Utilizing the hashtag hierarchy allows the analyst to explore the ranking results from a global overview to local details. Several filters, such as the edge or the glyph filter, enable analysts to customize this view easily. Relevant posts, hashtags, and users are also provided to help analysts better understand the content of the selected cluster node.

Smoothly switching between different data dimensions (R3). Inspired by the context popup interaction in [18], we also overlay context of a selected item to provide further navigation cues. For example, if the

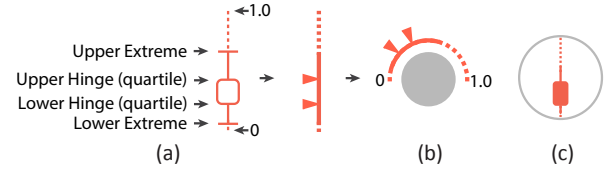


Fig. 7. Design of the uncertainty glyph: (a) box plot; (b) transforming the box plot to the uncertainty glyph; (c) an alternative design.

analyst selects a hashtag, the labels of users who use that hashtag can be overlaid around the selected hashtag via a word cloud (Fig. 11(a)). If the analyst finds something of interest, the hashtag graph will be smoothly transitioned to the user graph (Fig. 11(b)).

6.2 Uncertainty as Glyph

After testing with the first prototype, the experts identified several incorrect ranking results. They expressed the need to be informed of such results. This requirement is related intimately with the conclusion of previous work, which stated that effectively conveying uncertainty is very important to the visual analytics process [14, 49]. Since the ranking results are aggregated into clusters in the overview, the experts wanted to examine the uncertainty distribution of the aggregate node, including the minimum value (0), maximum value (1.0), lower extreme, upper extreme, lower hinge (25%), and upper hinge (75%).

Inspired by the box plot design (Fig. 7(a)), we have designed a glyph to meet the above requirements (Fig. 7(b)). As shown in Fig. 7(a), six values from a set of data are conventionally used in a box plot, including the minimum and maximum values, the extremes, and the upper and lower hinges (quartiles). A total of 50% percent of items fall in between the upper and lower hinges. To combine a box plot with a graph node, we first transform the box plot to a line-based one, and then bend it around the upper boundary of the node (Fig. 7(b)). We also attempted several alternatives in the participatory design process with experts. Fig. 7(c) is one of them. After interacting with this alternative, the experts stated that it was confusing. They thought that the item with more of a filled area inside should be the one on which they should focus. However, in reality, these nodes were only nodes with a larger area between the upper and lower hinges. A PhD student from an art school later confirmed that a larger amount of digital ink will attract more attention from users. After several interactions with the experts and the art student, we choose Fig. 7(b) as our final design.

Analysts can obtain an overview of the uncertainty distribution in a cluster by examining its uncertainty glyph. Fig. 8 illustrates several example patterns. For example, in Fig. 8(a), the majority of items in this cluster are characterized by low uncertainty. However, the cluster also contains some items with higher uncertainty. As a result, exploring the items with high uncertainty is a worthwhile endeavor.

Interaction. In addition to allowing analysts to examine the uncertainty score (R4), we also provide the interaction shown below to integrate an expert's knowledge into the retrieval process.

Interactive ranking refinement. After an expert finds an incorrect ranking result by examining the uncertainty glyph, the expert can modify the ranking result. The ranking scores of the corresponding graph nodes will also be updated accordingly. As shown in Figs. 1(c)- (f), the



Fig. 8. Four example patterns of the uncertainty glyph: (a) most items in this cluster have low uncertainty but some items with higher uncertainty are also included; (b) most items in this cluster have high uncertainty and some items with higher uncertainty also occur; (c) uncertainty distribution is uniform; (d) most items in this cluster have lower uncertainty.

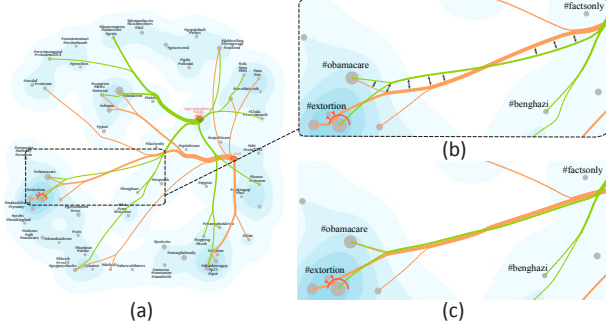


Fig. 9. Layout of multiple uncertainty propagation paths: (a) initial layout based on the flow map layout; (b) the matched result of the propagation paths; (c) the layout result of the propagation paths.

ranking scores (e.g., node sizes) of several nodes changed. A glyph is designed to illustrate the change, with the dotted orange circle encoding the previous ranking score and the boundary of the filled circle (gray color) representing the changed ranking score (Figs. 1(d)-(f)).

6.3 Uncertainty Propagation as Flow Map

The flow map [33, 44] is designed to visually analyze the movement of objects from one location to multiple locations. Inspired by this design, we develop the uncertainty propagation path (Fig. 1), which is useful for quickly deriving the unknown uncertain node(s) from the known one(s) (R5).

Layout. The layout of multiple uncertainty propagation paths of different nodes is based on the flow map layout in [44] and the edge bundling in [19]. The layout contains the following steps.

Step 1: Derive the initial uncertainty propagation path based on the flow map layout. We first compute the uncertainty propagation of the selected node based on the topology by using the method in Sec. 5.3. The flow map layout via spiral trees is then unitized to generate the initial uncertainty propagation path (Fig. 9(a)).

Step 2: Employ edge compatibility measures to match the corresponding propagation paths from different nodes. In this step, we employ the three compatibility measures described in [19] to match the propagation paths from different nodes.

The first measure is angle compatibility, which aims to match the edges with a smaller angle. It is defined by:

$$C_\alpha(e_i, e_j) = |\cos(\alpha)|.$$

The second measure is scale compatibility, which tends to match the edges with similar lengths. It is measured by:

$$C_s(e_i, e_j) = 2 / (l_{avg} \cdot \min(|e_i|, |e_j|) + \min(|e_i|, |e_j|) / l_{avg}),$$

where $l_{avg} = (|e_i| + |e_j|) / 2$.

The third measure is position compatibility, which aims to match the close edges together. It is defined by:

$$C_p(e_i, e_j) = l_{avg} / (l_{avg} + \|Q_{m1} - Q_{m2}\|),$$

where Q_{m1} and Q_{m2} are the midpoints of edges e_i and e_j .

The last measure, visibility compatibility, described in [19] is not considered in our method because there are too many line segments in the propagation path generated by the flow map layout, each of which

is quite short. Thus, if we consider this measure, many of these line segments will not be bundled together.

The total edge compatibility is defined by:

$$C_e(e_i, e_j) = C_\alpha(e_i, e_j) \cdot C_s(e_i, e_j) \cdot C_p(e_i, e_j).$$

Fig. 9(b) shows the matched results of the propagation paths.

Step 3: Compute the force to bundle the propagation path. The combined force for a point p_i on e_i is defined as:

$$F_{p_i} = K_i (\|p_{i-1} - p_i\| + \|p_i - p_{i+1}\|) + \sum_{e_j \in E} \|p_i - p_j\| \cdot C_e(e_i, e_j)$$

where K_i is the spring constant for each segment and E is the set of all the matched edges of e_i . In [19], the last item is the electrostatic force $F_e = 1 / \|e_i - e_j\|$. In order to bundle the matched paths that are located away from each other, we replace it with an attracting spring force. Fig. 9(c) shows the layout results of the propagation paths.

7 QUANTITATIVE EVALUATION

In this section, we quantitatively evaluate the effectiveness of our MRG computation and incremental ranking update algorithm.

7.1 MRG Computation

To evaluate the performance of our MRG computation based on the Monte Carlo sampling method, we compared it with the matrix-based method proposed in [16]. We used two Twitter datasets in the experiments: government shutdown and Ebola outbreak. The shutdown dataset contains tweets on the 2013 US government shutdown (5,132,510 tweets from Oct. 1 to Oct. 16, 2013), which were collected by using queries such as “shutdown.” The Ebola dataset contains tweets on the Ebola outbreak (1,425,017 tweets from Jan. 1 to Dec. 25, 2014), which were collected by using queries such as “ebola.” All experiments were conducted on a PC with a 3.1GHz CPU and 16 GB RAM.

There were too many posts, users, or hashtags and we could not label all of them. Thus, we did not report the recall in our evaluation. In this evaluation, we used top n-precision (n-Prec) as the evaluation measure. Top n-precision is the percentage of the correctly retrieved items among the top-n ranked items. This measure is often used when the recall is hard to calculate [9]. To fully compare the two algorithms, we calculated the top 10, 50, 100, and 200-precision for posts, users, and hashtags, respectively. We invited two PhD students who majored in data mining and are familiar with the datasets to evaluate the retrieval results. They labeled the results individually and resolved the differences via discussion. The results are shown in Table 1. Overall, our algorithm performed better than the baseline on both datasets. We inspected the top 10 retrieved items with both methods. In general, the retrieved items were quite accurate. However, the baseline had one mistake in the top 10 users selected from the shutdown dataset. It overestimated the importance of a user called @governmentcloud, who posted a significant number of tweets with a number of hashtags. However, this user did not have many followers and his/her tweets were seldom retweeted. In contrast, our algorithm can avoid this mistake by taking a user’s authority into consideration. The baseline algorithm also had similar mistakes in the Ebola dataset.

Dataset	n-Prec	Post		User		Hashtag	
		Base	Ours	Base	Ours	Base	Ours
Shutdown	10-Prec	1.000	1.000	0.900	1.000	1.000	1.000
	50-Prec	0.920	0.940	0.940	0.960	0.960	0.960
	100-Prec	0.870	0.930	0.850	0.920	0.910	0.930
	200-Prec	0.840	0.900	0.845	0.875	0.855	0.860
Ebola	10-Prec	1.000	1.000	0.800	1.000	1.000	1.000
	50-Prec	0.840	0.860	0.660	0.780	0.880	0.920
	100-Prec	0.770	0.840	0.660	0.730	0.870	0.880
	200-Prec	0.765	0.800	0.610	0.710	0.860	0.870

Table 1. Comparison of the MRG computation method with the baseline using top-n precision for posts, users, and hashtags.

7.2 Incremental Ranking Update

Since the incremental ranking update algorithm only calculates the statistics of the changed random walks, it is more efficient than the full update. In this section, we conducted an experiment to highlight the effectiveness of our incremental ranking update algorithm.

First, we demonstrate that the incremental algorithm converges quickly. To this end, we invited two analysts to use our system. One analyst worked on the Shutdown dataset while the other worked on the Ebola dataset. They updated the ranking incrementally based on the initial retrieval results. During the update process, when the analyst found that a ranking score of an item was underestimated, he increased its ranking score and vice versa. After each update, we re-calculated the top-200 precision for posts, users, and hashtags. After five updates, we observed that results were nearly unchanged from the last update. Hence, we allowed them to stop the process.

The results after each update are listed in Table 2. It shows that the retrieval results improved gradually as they interactively modified the ranking scores. This result verifies that our method can interactively refine the retrieval results by integrating analyst feedback.

We can further observe that after some updates, the performance of more than one type of item changed as well. For example, after changing the first item in the Ebola dataset, the performance of the retrieved posts, users, and hashtags all increased. This result confirmed the effectiveness of the MRG model and the developed computation method.

Update	Ebola			Shutdown		
	Post	User	Hashtag	Post	User	Hashtag
0	0.800	0.710	0.870	0.900	0.875	0.860
1	0.815	0.715	0.875	0.910	0.875	0.865
2	0.840	0.720	0.880	0.915	0.875	0.865
3	0.855	0.720	0.885	0.925	0.885	0.875
4	0.855	0.720	0.890	0.925	0.885	0.880
5	0.855	0.720	0.895	0.925	0.885	0.885

Table 2. Evaluation of our incremental ranking update algorithm. The changes are evaluated by top-200 precision. The row with 0 updates contains the initial results.

Second, since the incremental update algorithm can fully update the statistics of the changed random walks, the incremental update achieves the same ranking result as the full update algorithm.

8 APPLICATION

In order to evaluate the usefulness of MutualRanker, we performed two case studies on the same Twitter datasets described in Sec. 7. Due to the page limit, we focus our report on the shutdown dataset. Interested readers may refer to the attached video for the study on the Ebola dataset. Moreover, MutualRanker allows users to filter out irrelevant items based on their knowledge. For example, in the government shutdown case study, users can remove irrelevant hashtags such as “#retweet,” “#rt,” “#path,” and “#road” from the initial query.

The procedure of the case studies was loosely structured into three phases. First, we pre-interviewed two experts, one researcher in sociology (S) and one researcher in media and communications (C), to understand their respective interests in the datasets. We designed a number of exploration tasks. In the second phase, we collaborated with the experts to finish the designed tasks. During this phase, we asked questions to discuss with the experts the usefulness of our tool for each task. Finally, the experts were invited to another discussion session to provide overall feedback on how our tool could help them with real-world tasks.

8.1 Case Study: Government Shutdown

In this study, we worked with expert S to: 1) evaluate how uncertainty analysis can be utilized to identify key hashtags and users with a satisfactory confidence level; 2) leverage our system to iteratively reduce the uncertainty levels; 3) extract relevant hashtags/users/tweets related to the government shutdown.

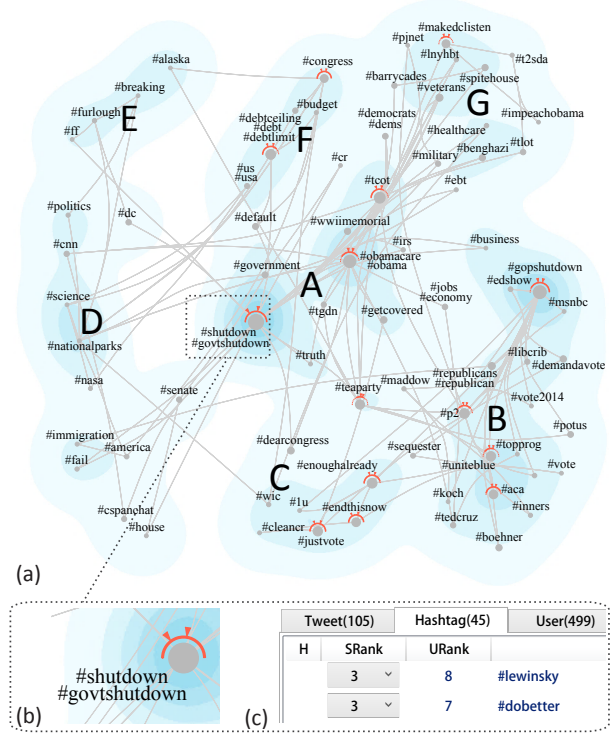


Fig. 10. The overview of the government shutdown dataset.

Overview. The expert quickly found interesting results after examining the hashtag overview (Fig. 10(a)) generated by our system. She identified seven prominent topics described by a set of hashtags: general discussions about the shutdown and Obamacare (Fig. 10A), political discourse on twitter (Fig. 10B), discussion on ending the shutdown (Fig. 10C), the influence of the shutdown on people’s lives (Fig. 10D) reporting the government shutdown on news media (Fig. 10E), debt-related discussion (Fig. 10F), and critics of the shutdown (Fig. 10G).

Uncertainty analysis: The “#shutdown” cluster (Fig. 10(b)) attracted the expert’s attention because it contains items with higher uncertainty. The expert examined the detailed hashtags and tweets in the cluster. She found that in addition to common hashtags such as #govtshutdown, #obamashutdown, and #shutdowngop, a number of diverse hashtags were also created. Such hashtags included those that criticized the shutdown, e.g., #shutdownharry; local news posts, e.g., #hounews, and public campaigns, e.g., #dontcutkids. She wanted to examine the most uncertain ones, so she sorted the hashtags by the uncertainty level. Interestingly, #lewinsky was ranked as the most uncertain hashtag (Fig. 10(c)). The analyst searched the related tweets and found that data tagged with #lewinsky concerned the shutdown of the Clinton government in 1995. The expert decided to lower the ranking score of the hashtag. During the process, she commented that the uncertainty glyph and item-filtering feature were useful, helping her filter out irrelevant items by lowering their ranking scores.

Uncertainty propagation: Next, the expert examined how the uncertainty of the “#shutdown” cluster would influence neighboring clusters. She clicked the “propagation” button and the corresponding uncertainty propagation was displayed (the orange flow in Fig. 1). She also selected the uncertainty propagation of the “#democrats” cluster (the blue flow in Fig. 1) and “#republicans” cluster (the green flow in Fig. 1), which were closely related to the “#shutdown” cluster. As shown in Fig. 1(b), cluster “#nationalparks” shared the uncertainty propagated from the three clusters. Given that the closing of the national parks was a result of the government shutdown and stimulated discussion on Twitter, the expert increased the ranking score of #nationalparks (from 4 to 6). In our system, the ranking score is from 1 to 10, with 10 being the highest score.

After the adjustment, she noticed the scores of another two hashtag clusters were automatically increased: “#spitehouse” and “#teaparty.” In the first cluster, the ranking scores of hashtags such as “#spitehouse”

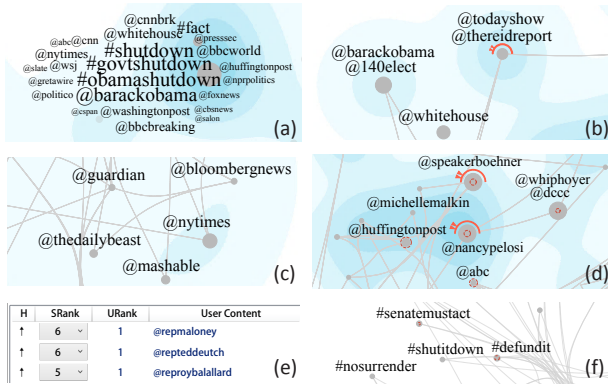


Fig. 11. Switching between different data graphs: (a) overlay labels around the focus cluster; (b) key government official accounts in the user graph; (c) news agencies/public media in the user graph; (d) the user graph after changing; (e) new ranking scores of a few politicians; (f) switch back to the hashtag graph.

and “#demshutdown” increased. In the second cluster, the ranking scores of hashtags such as “#teaparty” and “#defundgop” increased, as well. The expert commented, “It is helpful that the hidden relationships between hashtags are leveraged to propagate the ranking change. I can find more partisan messages around the topic and the public responses in this way.” She then found related tweets in the #spitehouse group and #teaparty group. For example, “@RepBradWenstrup @sarahlance #shutdown #Nationalpark Here’s what my tea-party-backed #Republican did to my vacation.”

On the contrary, the ranking score of cluster “#ebt” decreased, which is caused by the ranking score decrease of hashtags “#ebt” and “#obam-zombies.” The expert then examined the relevant tweets to probe the reason. The EBT system was crashed at that time and many people wondered whether the crash was caused by the government shutdown: “Ahh... #ebt not working cause if a #governmentshutdown? How sad you can’t spend money taken from me against my will that I worked for...” Then, the crash was explained to be a result of a computer failure (“According to NBC, #ebt is down because of a technical issue, NOT #governmentshutdown”). Thus, the expert believed “#ebt” was irrelevant and appreciated this automatic change.

Switching between different data views. In addition to hashtags, the expert wanted to examine the users who participated in different discussion groups. For example, she wanted to identify the most active users in the “#shutdown” cluster, so she overlaid the user labels around the hashtag labels (Fig. 11(a)). The expert then switched to the user view to explore additional user information (Fig. 11(b) and Fig. 11(c)). She immediately identified the leading users in Fig. 11(b) and Fig. 11(c). She described them with two categories: 1) key government official accounts, including “@barackobama,” “@whitehouse” (Fig. 11(b)); and 2) news agencies/public media such as “@nytimes,” “@guardian,” and “@bloombergnews” (Fig. 11(c)). Considering that partisan leaders were of major interest to her, she first observed the ranking scores of select politicians, e.g., @speakerboehner (Rank 8), @whiphoyer (Rank 8), @nancypelosi (Rank 7), etc. She believed that the importance of these user accounts was underestimated because the influence and activeness of politicians on twitter are usually much lower than that in real life. She changed the rankings of the partisan leaders, “@speakerboehner,” “@whiphoyer,” and “@nancypelosi,” to 10, which is the highest. Fig. 11(d) shows the difference after this refinement.

After the change, the user clusters were regenerated and the uncertainty levels of some nodes were largely reduced. Notably, “@whiphoyer” became an important cluster with the scores of several users in the cluster automatically increased (Fig. 11(e)). For example, “@repmaloney,” from 5 to 6 and “@repteddeutch,” from 5 to 6. “These are members of Congress. The change of their ranking scores is natural here.” The expert commented, “This is cool. [...] If I want to change the ranking score of one user, others just automatically follow. This could help me find the important users whose names I am not familiar with or who are not active on Twitter.”

The expert then switched back to the hashtag graph to check the influence of the change on this graph. She found a new hashtag cluster, “#senatemustact.” She then zoomed into this cluster. As shown in Fig. 11(f), the hashtag primarily expresses criticism of the government, blaming either the Democrats or Republicans (“@PeteSessions #DefundObamacare #shutdown #MakeDCListen #senatemustact Stand for the American People!”).

8.2 User Feedback

To evaluate the usefulness of our system, we conducted a semi-structured interview with the two experts. They used MutualRanker in the case study for 2 hours, so they were familiar with its basic functions. Overall, MutualRanker was well received by them.

The experts appreciated MutualRanker as a research tool to help them collect relevant posts, users, and hashtags quickly and conveniently. Expert C believed that MutualRanker is very useful for coding in media and communications. According to him, coding is the most labor-intensive work in his field. Extensive training and careful attention have always been required to produce reliable data. He commented, “A toolkit like MutualRanker is urgently needed in my daily work to reduce coding complexity and costs. Especially when there are not enough samples, the linkage between items [in this system] will provide more information to make decision. [...] This system also provides an opportunity to supervise the data retrieval process.”

Both the experts were impressed by the uncertainty illustration and its propagation function. For example, Expert C said, “Uncertainty propagation is an awesome feature, I can use it to find some unexpected data and increase the coverage of coding.”

The experts agreed that smoothly switching between different data graphs helped them find relevant data more quickly. Expert S commented, “This switching function enables me to easily transition between the hashtag graph and the user graph. When I modify one ranking score in one graph, I cannot only verify the result in this graph, but also verify it in another graph.”

The experts also suggested several improvements. The target audience of MutualRanker is experts with domain knowledge. The experts believed that average users can also benefit from it. They suggested that more intuitive visual design be used. Expert C said, “The uncertainty glyph can be simplified for a general user. For example, maybe the glyph does not need to encode the uncertainty distribution, just simply show that this ranking score is uncertainty.” They also expressed the need to retrieve streaming data.

9 DISCUSSION AND FUTURE WORK

This paper presents a visual analytics system, MutualRanker, to help analysts interactively retrieve data of interest from microblogs. We extend the MRG model to extract a multifaceted retrieval result that includes the mutual reinforcement ranking results, the uncertainty of each rank, and the uncertainty propagation among different graph nodes. The model is tightly integrated with a composite visualization to assist analysts in retrieving salient posts, users, and hashtags effectively, in an uncertainty-aware environment.

In the future, we plan to improve system performance by implementing a parallel Monte Carlo sampling method. Another exciting avenue for future work is to retrieve streaming data in microblogs, which can be very useful in emergency management and threat analysis. We believe the system can also benefit average users interested in collecting microblog data. In the future, we will also invite more users to try our system and conduct a formal user study. Accordingly, we will improve MutualRanker based on the collected feedback.

ACKNOWLEDGMENTS

We would like to thank X. Wang and J. Yin, J. Gong, and Dr. W. Cui for helpful discussions on the visualization design, Dr. J. Zhang and Dr. Y. Song for constructive suggestions on similarity measures, as well as Dr. W. Peng and Dr. J. Su for providing domain expertise.

REFERENCES

- [1] A. C. Alhadi, T. Gottron, J. Kunegis, and N. Naveed. Livetweet: Microblog retrieval based on interestingness and an adaptation of the vector space model. In *Proceedings of TREC*, 2011.
- [2] K. Avrachenkov, N. Litvak, D. Nemirovsky, and N. Osipova. Monte carlo methods in pagerank computation: When one iteration is sufficient. *SIAM J. Numer. Anal.*, 45(2):890–904, 2007.
- [3] B. Bahmani, A. Chowdhury, and A. Goel. Fast incremental and personalized pagerank. *Proc. VLDB Endow.*, 4(3):173–184, 2010.
- [4] M. Bianchini, M. Gori, and F. Scarselli. Inside pagerank. *ACM Trans. Internet Technol.*, 5(1):92–128, 2005.
- [5] I. BIPM, I. IFcc, and I. IuPac. Oiml, guide to the expression of uncertainty in measurement. *International Organization for Standardization, Geneva. ISBN*, pages 92–67, 1995.
- [6] H. Bosch, D. Thom, F. Heimerl, E. Püttmann, S. Koch, R. Krüger, M. Wörner, and T. Ertl. Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering. *IEEE TVCG*, 19(12):2022–2031, 2013.
- [7] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.
- [8] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE TVCG*, 18(12):2649–2658, 2012.
- [9] A. Chandramouli and S. Gauch. A co-operative web services paradigm for supporting crawlers. In *Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*, pages 475–489, 2007.
- [10] H. Chen, S. Zhang, W. Chen, H. Mei, J. Zhang, A. Mercer, R. Liang, and H. Qu. Uncertainty-aware multidimensional ensemble data visualization and exploration. *IEEE TVCG*, 2015 (To Appear).
- [11] J. Chen, J. Zhu, Z. Wang, X. Zheng, and B. Zhang. Scalable inference for logistic-normal topic models. In *Proceedings of NIPS*, pages 2445–2453, 2013.
- [12] S. Cherichi and R. Faiz. Relevant information management in microblogs. *Information Systems for Knowledge Management*, pages 159–182, 2013.
- [13] C. Collins, S. Carpendale, and G. Penn. Visualization of uncertainty in lattices to support decision-making. In *Proceedings of EUROVIS*, pages 51–58, 2007.
- [14] C. Correa, Y.-H. Chan, and K.-L. Ma. A framework for uncertainty-aware visual analytics. In *Proceedings of IEEE VAST*, pages 51–58, Oct 2009.
- [15] D. R. Cox and P. A. Lewis. *The statistical analysis of series of events*. Wiley, 1966.
- [16] Y. Duan, F. Wei, Z. Chen, M. Zhou, and H. Shum. Twitter topic summarization by ranking tweets using social influence and content quality. In *Proceedings of Coling*, pages 763–780, 2012.
- [17] M. Efron. Information search and retrieval in microblogs. *Journal of the American Society for Information Science and Technology*, 62(6):996–1008, 2011.
- [18] S. Ghani, B. Kwon, S. Lee, J.-S. Yi, and N. Elmqvist. Visual analytics for multimodal social network analysis: A design study with social scientists. *IEEE TVCG*, 19(12):2032–2041, 2013.
- [19] D. Holten and J. J. Van Wijk. Force-directed edge bundling for graph visualization. *Computer Graphics Forum*, 28(3):983–990, 2009.
- [20] W. Javed and N. Elmqvist. Exploring the design space of composite visualization. In *Proceedings of PacificVis*, pages 1–8, 2012.
- [21] T. Kamada and S. Kawai. An algorithm for drawing general undirected graphs. *Information processing letters*, 31(1):7–15, 1989.
- [22] O. D. Lampe and H. Hauser. Interactive visualization of streaming data with kernel density estimation. In *Proceedings of PacificVis*, pages 171–178, 2011.
- [23] B. Liu and L. Zhang. A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463, 2012.
- [24] S. Liu, W. Cui, Y. Wu, and M. Liu. A survey on information visualization: recent advances and challenges. *The Visual Computer*, pages 1–21, 2014.
- [25] S. Liu, X. Wang, J. Chen, J. Zhu, and B. Guo. Topicpanorama: A full picture of relevant topics. In *Proceedings of IEEE VAST*, pages 183–192, 2014.
- [26] X. Liu, Y. Song, S. Liu, and H. Wang. Automatic taxonomy construction from keywords. In *Proceedings of KDD*, pages 1433–1441, 2012.
- [27] S. Lodha, A. Pang, R. Sheehan, and C. Wittenbrink. Uflow: visualizing uncertainty in fluid flow. In *Proceedings of IEEE Visualization*, pages 249–254, Oct 1996.
- [28] Y. Lu, F. Wang, and R. Maciejewski. Business intelligence from social media: A study from the vast box office challenge. *IEEE Computer Graphics and Applications*, 34(5):58–69, 2014.
- [29] Z. Luo, M. Osborne, S. Petrovic, and T. Wang. Improving twitter retrieval by exploiting structural information. In *Proceedings of AAAI*, 2012.
- [30] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller. Twinfo: Aggregating and visualizing microblogs for event exploration. In *Proceedings of CHI*, pages 227–236, 2011.
- [31] R. McCreadie and C. Macdonald. Relevance in microblogs: Enhancing tweet retrieval using hyperlinked documents. In *Proceedings of OAIR*, pages 189–196, 2013.
- [32] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, 1997.
- [33] D. Phan, L. Xiao, R. Yeh, and P. Hanrahan. Flow map layout. In *Proceedings of IEEE InfoVis*, pages 219–224, 2005.
- [34] E. J. Ruiz, V. Hristidis, C. Castillo, A. Gionis, and A. Jaimes. Correlating financial time series with micro-blogging activity. In *Proceedings of WSDM*, pages 513–522, 2012.
- [35] S. Sedhai and A. Sun. Hashtag recommendation for hyperlinked tweets. In *Proceedings of SIGIR*, pages 831–834, 2014.
- [36] L. Shi, F. Wei, S. Liu, L. Tan, X. Lian, and M. Zhou. Understanding text corpora with multiple facets. In *Proceedings of IEEE VAST*, pages 99–106, 2010.
- [37] M. Skeels, B. Lee, G. Smith, and G. G. Robertson. Revealing uncertainty for information visualization. *Information Visualization*, 9(1):70–81, 2010.
- [38] A. Slingsby, J. Dykes, and J. Wood. Exploring uncertainty in geodemographics with interactive graphics. *IEEE TVCG*, 17(12):2545–2554, 2011.
- [39] G. Sun, Y. Wu, R. Liang, and S. Liu. A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology*, 28(5):852–867, 2013.
- [40] G. Sun, Y. Wu, S. Liu, T.-Q. Peng, J. Zhu, and R. Liang. Evoriver: Visual analysis of topic competition on social media. *IEEE TVCG*, 20(12):1753–1762, 2014.
- [41] J. Tang, Z. Liu, M. Sun, and J. Liu. Portraying user life status from microblogging posts. *Tsinghua Science and Technology*, 18(2):182–195, 2013.
- [42] J. Thomson, E. Hetzler, A. MacEachern, M. Gahegan, and M. Pavel. A typology for visualizing uncertainty. *SPIE*, 5669:146–157, 2005.
- [43] C. Vehlouw, T. Reinhardt, and D. Weiskopf. Visualizing fuzzy overlapping communities in networks. *IEEE TVCG*, 19(12):2486–2495, Dec 2013.
- [44] K. Verbeek, K. Buchin, and B. Speckmann. Flow map layout via spiral trees. *IEEE TVCG*, 17(12):2536–2544, 2011.
- [45] F. Wei, W. Li, Q. Lu, and Y. He. Query-sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization. In *Proceedings of SIGIR*, pages 283–290, 2008.
- [46] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: Finding topic-sensitive influential twitterers. In *Proceedings of WSDM*, pages 261–270, 2010.
- [47] Y. Wu, S. Liu, K. Yan, M. Liu, and F. Wu. Opinionflow: Visual analysis of opinion diffusion on social media. *IEEE TVCG*, 20(12):1763–1772, 2014.
- [48] Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu. Opinion-seer: Interactive visualization of hotel customer feedback. *IEEE TVCG*, 16(6):1109–1118, 2010.
- [49] Y. Wu, G.-X. Yuan, and K.-L. Ma. Visualizing flow of uncertainty through analytical processes. *IEEE TVCG*, 18(12):2526–2535, Dec 2012.
- [50] P. Xu, Y. Wu, E. Wei, T.-Q. Peng, S. Liu, J. J. H. Zhu, and H. Qu. Visual analysis of topic competition on social media. *IEEE TVCG*, 19(12):2012–2021, 2013.
- [51] E. Zangerle, W. Gassler, and G. Specht. On the impact of text similarity functions on hashtag recommendations in microblogging environments. *Social Network Analysis and Mining*, 3(4):889–898, 2013.
- [52] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, and C. Collins. #fluxflow: Visual analysis of anomalous information spreading on social media. *IEEE TVCG*, 20(12):1773–1782, 2014.
- [53] X. W. Zhao, Y. Guo, Y. He, H. Jiang, Y. Wu, and X. Li. We know what you want to buy: A demographic-based system for product recommendation on microblogs. In *Proceedings of KDD*, pages 1935–1944, 2014.
- [54] H.-J. Zimmermann. *Fuzzy set theory and its applications*. Springer Science & Business Media, 2001.
- [55] T. Zuk and S. Carpendale. Visualization of uncertainty and reasoning. In *Proceedings of SG*, pages 164–177, 2007.